UDC 519.17:512.5 DOI https://doi.org/10.26661/2414-0287-2019-2-42-13

THE USAGE OF EVOLUTIONARY ALGORITHMS FOR SEARCHING OPTIMAL CLASSIFICATIONS

Kozin I.V., Maksyshko N.K., Selyutin E.K.

Zaporizhzhia National University
Ukraine, 69600, Zaporizhzhia, Zhukovsky str., 66

ainc00@gmail.com, wizz92@gmail.com

Key words:
classification, evolutionary algorithm, genetic algorithm, crossing, mutation, selection.

The article discusses the task of classification, the need to fulfil conditions and relations on the set. The relevance of the fragmentary structure of the task has been emphasized. It is illustrated an example of a classification problem that is considered in this paper and the task of covering stars with stars that occurs in many economic applications. As elementary fragments all edges appear in the graph. The conditions for joining an edge – this edge is the beam of an already existing star or has no common vertices with already built-up stars of coverage. The emphasis is on the lack of optimality of such a solution. The relevance of the fragmentary structure of the task has been emphasized. The possibility of constructing classes has been given, considering the entire list of objects classified in a certain sequence. Based on the fragmentary structure, it has been proposed to use an evolutionary algorithm. The work has analysed the examples of the practical application of the evolutionary (genetic) algorithm for solving the classification problems. The prospect of using a genetic algorithm for finding optimal classifications has been estimated. The following is a step-by-step sequence of operations of the genetic algorithm with examples: selection, crossing, mutation, selection. Examples of key operators, namely crossover and mutations, have been given. The detailed algorithm of the evolutionary model has been clearly illustrated. The principle of the evolutionary-fragmentary algorithm has been written in detail. Like a set of admissible solutions, a subset of maximal fragments on a given fragmentary structure has been considered. The mechanism of testing the quality of the genetic algorithm on a fragmentary structure, which is reduced to the enumeration of many variants, has been determined.

ВИКОРИСТАННЯ ЕВОЛЮЦІЙНИХ АЛГОРИТМІВ ПОШУКУ ОПТИМАЛЬНИХ КЛАСИФІКАЦІЙ

Козін І. В., Максишко Н.К., Селютін Є.К.

Запорізький національний університет
Україна, 69600, м. Запоріжжя, вул. Жуковського, 66

Ключові слова:
класифікація, еволюційний алгоритм, генетичний алгоритм, схрещування, мутація, селекція.

Розглянута задача класифікації, необхідність виконання умов й відносин на множині. Указані властивості відношення еквівалентності. Наведений приклад задачі класифікації, який буде розглядатися в роботі, а саме задача покриття графа зірками, яка виникає в багатьох економічних додатках. Як елементарні фрагменти виступають усі ребра графа. Умови приєднання ребра –ребро є променем існуючої зірки або не має спільних вершин із уже побудованими зірками покриття. Акцентована увага на недостатній оптимальності такого рішення. Підкреслена актуальність фрагментарної структури задачі. Указана можливість побудови класів, розглядаючи весь список об’єктів, що класифікуються в певній послідовності. На базі фрагментарної структури запропоновано використовувати еволюційний алгоритм. Проаналізовані роботи, у яких розглядаються приклади практичного застосування еволюційного (генетичного) алгоритму для розв'язування задач класифікації. Оцінена перспектива використання генетичного алгоритму пошуку оптимальних класифікацій. Указана поетапна послідовність операцій генетичного алгоритму з прикладами: відбір, схрещування, мутація, селекція. Наведені приклади роботи ключових операторів, а саме кросовера та мутації. Наочно проілюстрований детальний алгоритм еволюційної моделі. Детально описано принцип дії еволюційно-фрагментарного алгоритму. Як множина допустимих розв'язків розглядається підмножина максимальних фрагментів на заданій фрагментарній структурі. Визначений механізм перевірки якості генетичного алгоритму на фрагментарній структурі, який зводиться до перебору багатьох варіантів.

Statement of the problem

The task of classification arises in virtually all branches of knowledge: analysing research results; designing and forecasting; evaluating and making decisions. Despite the seeming simplicity of the problem, the task of classification can be very complex and controversial. In such cases, metaheuristics of different types are used to solve the problem [1].

Effective is the approach to the problem of classification based on the theory of fragmentary structures [2]. Like any other metaheuristic, the fragmentary approach is inferior to the classical approximate algorithms and cannot guarantee the proximity of the solution, which is found with its help, to the optimal one. However, the simplicity of implementation and the high convergence rate make it possible to find good approximate solutions of many complex optimization problems using this method [3-5].

This study discusses the use of a combination of fragmented and evolutionary algorithms for searching optimal classifications.

Analysis of recent studies and publications

Currently, there are a large number of various classification methods, the effectiveness of which essentially depends on the specifics of the subject area in which this task is formulated and the characteristics of the initial information. In the works of A.Zenkina, V.Vapnik, Yu.Zhuravlev, examples of the theoretical application of genetic algorithms for solving classification problems have been considered [6–8].

The works of such scientists as S.Ayvazyan, M.Aizerman, N.Zagoruiko, V.Mazurov, L.Rastrigin are devoted to the practical application of genetic algorithms [9-13]. D.Goldberg has proposed the use of a crossing operator [14].

A number of approaches to the classification problem have been considered [15].

The basic idea of finding a “good” classification is to maintain an optimality criterion. In this case, the classification problem is reduced to the problem of finding the optimal solution on some set of feasible solutions. And then all sorts of optimization methods start to work.

Objectives of the article

The purpose of the article is to assess the prospects for the use of a genetic algorithm for finding optimal classifications.

The main material of the research

The classification problem on the set X is the task of partitioning the set X into disjoint classes. The task of finding such a family of subsets of the set X for which the following conditions have been completed:

$$X_\alpha \subseteq X, X_\alpha \neq \varnothing$$

$$\bigcap_\alpha X_\alpha = X$$

The requirement of non-emptiness of classes is immaterial. We will consider the same such classifications.

Recall that the equivalence relation on the set X is any binary reflexive, symmetric, and transitive relation on this set, that is, the relation “~”, which is determined by the properties:

$$\forall x \in X\ x \approx x$$

$$\forall x, y \in X\ x \approx y \Rightarrow y \approx x$$

$$\forall x, y, z \in X\ x \approx y, y \approx z \Rightarrow x \approx z.$$

Each classification generates a natural equivalence relation on the set X. Namely: two elements are equivalent only if they belong to the same class.

An example of a classification problem that will be considered in this paper is the problem of covering a graph with stars, which arises in many economic applications [16].

For example, the given graph has edges weighted by non-negative numbers. The task is to find the set of stars in this graph, which contain all the vertices of the graph and the total weight of the edges, which is minimal. This task is actually a classification task. The class, in this case, is all the top of the star. In each class a representative has been selected – the centre of the star.

Stars are ordered in an arbitrary way. At the next step of the algorithm, the first star in the sequence of stars is selected, which does not have common vertices with the fragments which have been already found.

The idea of a fragmentary algorithm is as follows: each feasible solution to the problem is a fragment that is represented as the union of indivisible components – elementary fragments. On the set of fragments, the join operation is introduced, which allows to obtain a new fragment by adding one of the elementary fragments to the existing one. A fragmentary algorithm is an algorithm for finding the maximum fragment by inclusion.

In details:

– The set of solution fragments {f} and the external operation $\oplus$ joining fragments is determined. Each admissible solution to the problem consists of a finite number of fragments from a given set of fragments connected by an external operation.

– A series of relations of a linear order {} on a set of fragments and an effective procedure for ordering fragments, which allows to obtain the ordering of fragments in different orders in accordance with the selected rules, is specified.

– Conditions for the attachment of a fragment are specified, which can be both deterministic and dynamic, changing at each step. For an already selected subset of fragments, an effective procedure is established to verify the possibility of attaching a fragment that does not belong to the selected subset.

Thus, a fragmentary algorithm allows us to construct a feasible solution of the problem in a time linear in the number of fragments.

We show that the problem of covering a graph with stars has a fragmentary structure and, accordingly, any feasible solution to the problem can be obtained by applying a fragmentary algorithm.

The elementary fragments are all the edges of the graph. An edge attachment condition – this edge is a ray of an already existing star or has no common vertices with the already constructed covering the stars. It is easy to show that by applying a fragmentary algorithm with a certain choice of a sequence of edges, you can get any set of stars in the graph, that is, any feasible solution to the classification problem. However, such solutions are not necessarily optimal. In order to find the optimal solution, we describe a combination of evolutionary and fragmentary algorithms.

Evolutionary (genetic) algorithms have been considered in detail in numerous publications [15,16]. Genetic algorithms are an optimization method based on the evolution of an "individual" population.

In 1975, G. Holland's book "Adaptation in natural and artificial systems" was published, in which the genetic algorithm was proposed [16].

For a number of optimization problems, it was possible to propose quite effective procedures for finding optimal solutions based on the use of evolutionary algorithms. To implement an evolutionary algorithm, it is necessary to select a number of objects and procedures, the totality of which will be called the evolutionary model [15]. The main components of the evolutionary model are the following:

- the basic solution set, the set of feasible solutions X, on which the optimal solution of the problem is sought;
- the operator of constructing the initial population: a procedure that allows you to select on the set of all feasible solutions its subset Y⊆X for subsequent evolution;
- the selection criterion, an algorithm that allows you to compare the quality of solutions within a given population;
- the crossover operator $K : X \times X \to X$, which allows for two acceptable decisions-parents to build a new solution, a descendant of a set of valid solutions;
- the mutation operator $M : X \to X$;
- the selection operator, which allocates a set of pairs in Y to perform the operation of the crossover;
- the evolution operator, allowing to build new populations from a variety of parents and descendants;
- the stop rule, which defines the condition for stopping the evolutionary algorithm.

We describe briefly the principle of the evolutionary algorithm. At the initial step, the set of solutions Y0 is constructed using the initial population operator. At each next step, it is assumed to be given a certain set of permissible solutions – the current population. The first step is a lot $Y = Y_0$. For each element of the set Y, the value of the selection criterion is calculated. Then, using the selection operator in the current population Y, a set of pairs for the crossover is selected. A crossover operator is applied to each pair of the selected set of pairs, and then a mutation operator is applied to the result of the crossover. In this way there are many elements – descendants $\tilde{Y}$.

For the intermediate population $Y \cup \tilde{Y}$, which is the union of the current population and the set of descendants, the evolution operator is used, which highlights the new current population on this set. The process of evolution is repeated until the condition for stopping the evolutionary algorithm is fulfilled. The block diagram of the evolutionary algorithm is shown in Figure 1.

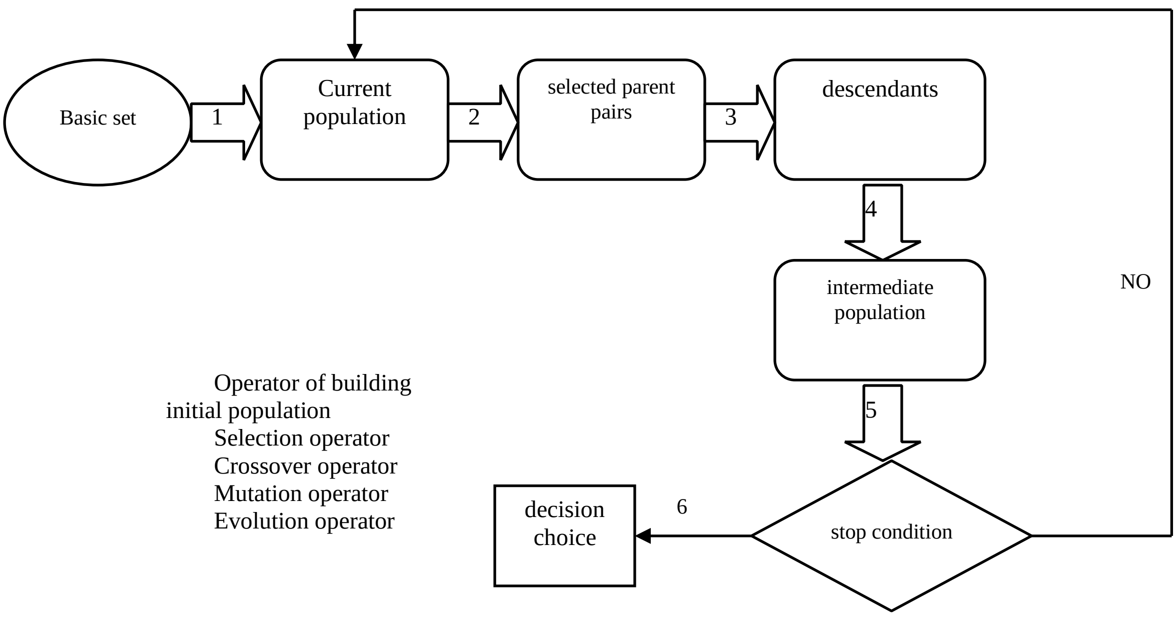

Fig. 1. Evolutionary model

The properties of fragmented structures allow us to construct a special class of evolutionary algorithms on fragmented structures – EVF-algorithms.

EVF – the algorithm is a combination of the evolutionary and fragmentary algorithm. We describe the evolutionary model and principle of operation of such algorithm.

A subset of maximal fragments on a given fragmentary structure is considered as a set of feasible solutions. Each fragment of this set is defined as the result of the work of the fragmentary algorithm with any given permutation of elementary fragments. Thus, any permissible solution corresponds to a certain permutation of the numbers 1,2, ..., N, where N is the number of elementary fragments. For each valid solution, the value of the objective function is determined.

The base set X of the evolutionary model is the set $S_N = \{i_1, i_2, ..., i_N\}$ of all permutations of the numbers 1,2, ..., N. The initial population construction operator selects a subset of the given power Q from the set X.

The rule for calculating the selection criterion is arranged as follows: for a given permutation of fragments, using the fragmentary algorithm, the maximum allowable fragment is constructed. The value of the objective function of the problem for this fragment is calculated.

We now describe the crossover operator. Let $U = (u_1, u_2, ..., u_N)$ and $V = (v_1, v_2, ..., v_N)$ be two arbitrary permutations. The permutation — the descendant is constructed as follows: the sequences U and V are viewed from the beginning. At the *k*-th step, the smallest of the first sequence elements is selected and added to the new permutation – a descendant. This element is then removed from the two parents-sequences. For example,

K ((2,3,4,7,8,1,6,5),(3,4,6,2,1,5,8,7)) = (2,3,4,6,1,5,7,8)

The mutation operator M performs a random transposition (replacing two elements with places) in a permutation.

The selection operator randomly selects a set of pairs from a given number of pairs in the set of permutations of the current population.

The evolution operator arranges the elements of the intermediate population into a sequence in descending order of the value of the selection criterion. The first Q elements of the sequence are selected as the new current population.

The usual rule of stopping is that the number of generations has reached the limit of L. The best permutation criterion permutation from the last constructed population determines the approximate solution of the problem.

The genetic algorithm on a fragmentary structure is heuristic. In fact, it is limited to sorting through the many options. It is almost impossible to evaluate the quality of such an algorithm in the absence of exact solution algorithms. It is only reasonable to compare the operation of this algorithm with other approximate algorithms on large series of problems.

The following mechanism is proposed for checking the quality of the fragmentary-genetic algorithm for a certain class of problems.

It randomly builds a sequence of tasks of this class.

For each task of this sequence, a solution is sought in two ways:

1. ordering of fragments occurs through the use of genetic algorithms;

2. ordering of fragments is performed randomly.

The number of draws in the second case is equal to the size of the population size generated by applying the genetic algorithm in the first case.

The quality of a genetic algorithm is determined by the ratio of the number of tasks for which the solution obtained using the genetic algorithm is better than the solution obtained by random search to the total number of tasks in the series.

The proposed approach is universal and allows applying the same evolutionary algorithm to any optimization problems on finite fragmentary structures.

Conclusions

The evolutionary algorithm reflects the synthesis of algorithms for classifications and optimization of the functional characterizing the quality of classification.

The method is focused on processing multidimensional arrays of information, the features of which are the high dimensionality of the attribute space and the small sample size of objects. The proposed classification method makes it possible not to preliminarily reduce the dimensionality of the attribute space, which, in turn, makes it possible to exclude the loss of significant information and to take into account internal relations in the information arrays under consideration.

The results allow us to conclude that the approach to finding the optimal solution to a number of discrete optimization problems based on the use of a combination of fragmentary and genetic algorithms is promising.

This approach is especially useful in decision support systems, when the final question about the quality of a decision is decided on the basis of the decision-maker's decision.

References

1. Skobtsov Yu.A. (2008). *Osnovy evolutsionyh vychislenij [Fundamentals of evolutionary computing]*. Donetsk: DonNTU [in Russian].

2. Kozin I. V., Maksyshko N.K. & Perepelitsa V.O. (2017). Fragmentary Structures in Discrete Optimization Problems. *Cybernetics and Systems Analysis November, 53*, 931–936.

3. Kozin I.V., Batovskiy S.E., & Sardak V.I. (2017). Fragmentarna model i evolutsijnii algorytm dlya 2d upakovki obyektiv [Fragmentary model and evolutionary algorithm for 2d object packaging]. *Sat. Mathematical and computer model. Physico-mathematical sciences, 15*, 74–79 [in Ukrainian].

4. Kozin I.V., Krivtsun E.V, & Pinchuk V.P. (2015). Evolutsijno-fragmentarna model zadachi trassirovki [Evolutionary fragmentary model of the trace problem]. *Cybernetics and Systems Analysis, 3,* 35–50 [in Ukrainian].

5. Kozin I.V., & Polyuga S.I. (2014). Fragmentarni modeli dlya deyakyh ekstremalnyh zadach na grafah [Fragmentary models for some extremal problems on graphs]. *Mathematical machines and systems, 1,* 143–150 [in Ukrainian].

6. Zenkin A.A. & Zenkin A.I. (2010) Zadacha postroenia optimalnyh klassifikatsij [The task of building optimal classifications]. *Collection of papers on the mat. Cybernetics EC of the USSR*, 20—33 [in Russian].

7. Vapnik V.N., & Chervonenkis A.Ya. (1974). *Teorija rozpoznanya obrazov (statistichni problemi) [Theory of pattern recognition (statistical learning problems)].* M.: Science [in Russian].

8. Zhuravlev Yu. I. (1998). *Izbranny nauchni trudi [Selected Scientific Works]*. M: Publishing Master [in Russian].

9. Ayvazyan S. A., & Buchstaber V.M. (1989). *Prykladna statistika: klassyfikatsia i znizhenya rozmirnosti [Applied statistics: Classification and reduction of dimension]*. M.: Finance and Statistics [in Russian].

10. Aizerman M. A., & Bravermann E.M. (1970). *Metod potentsijnyh funksii v teorii obuchenya mashin [The method of potential functions in the theory of learning machines]*. M .: Science [in Russian].

11. Zagoruiko N. G. (1999). *Prykladnye metody analiza dannyh i znanyi [Applied methods of data and knowledge analysis]*. Novosibirsk: Publishing house of the Institute of mathematics [in Russian].

12. Mazurov V.D. (1990). *Metod komitetov v zadachah optimizatsii i klasifikatsii [The method of committees in the tasks of optimization and classification]*. M. : Science [in Russian].

13. Rastrigin LA, & Erenstein R.H. (1981). *Metod kolektivnogo raspoznavanya [The method of collective recognition]*. M .: Energoizdat [in Russian].

14. Goldberg D.E. (1989) *Genetic algorithms in search, optimization, and machine learning*. Addison Wesley Longman.

15. Perepelitsa V.O., Kozin I.V., & Tereshchenko I.V. (2008) *Zadachi klasifikatsii: pidhody, metodi, algorytmi [Classification tasks: approaches, methods, algorithms]*. Poligraf, Zaporizhzhya [in Ukrainian].

16. Maksishko N. K., & Zakhovalko T.V. (2009) *Modely ta metody rozvyazannya prykladnih zadach pokruttya na grafah i gipergrafah [A model of methods and application of applied problems of covering on graphs and hypergraphs]*. Poligraf, Zaporizhzhya [in Ukrainian].

17. Holland J. H. (1975) *Adaptation in natural and artificial systems*. Ann Arbor. MI: University of Michigan Press.